

Calculating transition rates from durations of transition paths

Fangqiang Zhu^{a)}

Department of Physics, Indiana University - Purdue University Indianapolis, Indianapolis, Indiana 46202, USA

(Received 24 January 2017; accepted 6 March 2017; published online 29 March 2017)

Calculating the kinetic rates for rare transitions is an important objective for molecular simulations. Here I prove equalities that relate the transition rates to the equilibrium free energy and the statistics of the transition paths. In particular, the durations of the transition paths within given intervals of the reaction coordinate provide the kinetic pre-factor in the rate formula. Based on the available free energy, the transition rates can further be rigorously calculated by initiating forward and backward simulations and evaluating the duration of each transition path. Validation on a model system confirms that the approach correctly predicts the transition rates from the simulations and demonstrates that whereas the relations here are general and valid for any chosen reaction coordinate, a good reaction coordinate will enable a more efficient sampling of the transition paths and thus a more reliable rate calculation. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4979058>]

I. INTRODUCTION

Rare transitions¹ between metastable states, such as chemical reactions or conformational changes of biomolecules, are often of major interest in physics, chemistry, and biology. The forward and backward transition rates are the major kinetic quantities for any two-state transition, with their ratio also determining the equilibrium probabilities of the metastable states. For molecular systems, all-atom molecular dynamics (MD) simulation is potentially a powerful tool to reveal the ultimate details of the transitions. However, because in equilibrium the system would stay in a metastable state for a very long time before making a rare transition, the affordable sampling times of straightforward MD simulations are often far from enough to produce spontaneous transitions and to directly obtain the transition rates.

To circumvent the time scale problem, a variety of computational approaches have been designed to calculate the transition rates from MD simulations. In one important category of such approaches, transition path sampling^{1,2} and its variants such as transition interface sampling^{3,4} calculate the kinetic rates from the ensemble of transition paths. Alternatively, it is also common to employ a reaction coordinate⁵ to facilitate the rate calculations. Typically, a free energy along the reaction coordinate is first obtained using some enhanced sampling techniques,⁶ and then the transition rates are further calculated based on the free energy. For example, if the time evolution of the system can be described as a diffusion process along the reaction coordinate,⁷ the Smoluchowski equation can be established with the estimated diffusion coefficients⁸ to calculate the transition rates.^{9,10} If the time evolution of the reaction coordinate can be assumed to be a Markov process, some methods^{11–13} could provide the transition rates from the same simulations for the free energy calculation. Under the assumption of a perfect reaction coordinate and ideal barrier crossing,

the transition state theory^{1,14} estimates the kinetic rates from the free energy and the average speed at the barrier top. In general, the rate in the transition state theory differs from the exact rate by a factor termed as transmission coefficient,^{1,14} and Bennett and Chandler formulated^{1,14,15} methods to calculate the transmission coefficient from unbiased simulations with the system released at the barrier top.

By relating the equilibrium and transition-path ensembles, Hummer proposed new theories^{16,17} for calculating the transition rates. One particularly appealing method calculates the rates from unbiased trajectories launched at the free energy barrier,^{16,17} similar to the Bennett-Chandler approach above but with superior statistical efficiency, especially when the transition paths re-cross the barrier multiple times. Following similar ideas, Daru and Stirling recently proposed a method¹⁸ to obtain the rates from the residence time of the transition paths in the “saddle domain” near the free energy barrier. In this article, I will present rigorous derivations for the exact relations between the transition rate and the durations of the transition paths, thus making it possible to calculate the kinetic rate from MD simulations without approximations. I will also use a model system to validate the approach and to demonstrate the effect of the reaction coordinate on the rate calculation.

II. THEORY

Consider a classical system with two alternative metastable states, A and B . Assume that a reaction coordinate α has been defined, such that any given microstate (i.e., the positions and momenta of all atoms) of the system can be mapped to a single value of α . The thermodynamics of the system can then be represented by the normalized equilibrium probability distribution $p_\alpha(\alpha)$ or, equivalently, a one-dimensional free energy $G(\alpha) \equiv -k_B T \cdot \ln p_\alpha(\alpha) + \text{const.}$, with k_B the Boltzmann constant and T the temperature. The equilibrium distribution $p_\alpha(\alpha)$ or the free energy $G(\alpha)$ can be calculated using standard enhanced sampling techniques⁶ such as umbrella sampling simulations.^{10,19} For a two-state system, $G(\alpha)$ typically features two major minima separated by a barrier. Here I use α_A ,

^{a)} Author to whom correspondence should be addressed. Electronic mail: fzhu0@iupui.edu

α_B , and α^* to denote the locations of the minima and the barrier, respectively, with $\alpha_A < \alpha^* < \alpha_B$. The probability distribution $p_\alpha(\alpha)$ thus has two peaks at α_A and α_B , respectively, and the integration of each peak represents the equilibrium probability for the corresponding metastable state,

$$P_A = \int_{-\infty}^{\alpha^*} p_\alpha(\alpha) d\alpha, \quad (1a)$$

$$P_B = \int_{\alpha^*}^{\infty} p_\alpha(\alpha) d\alpha. \quad (1b)$$

Kinetically, I use k_0 to denote the rate of spontaneous transitions from state A to B (i.e., the probability of observing such a transition in unit time) in the equilibrium ensemble, which is also equal to the rate of spontaneous transitions from B to A . When k_0 is known, the forward and backward transition rates are given by

$$k_{A \rightarrow B} = k_0 / P_A, \quad (2a)$$

$$k_{B \rightarrow A} = k_0 / P_B. \quad (2b)$$

Therefore, the rate calculation amounts to obtaining k_0 .

The time evolution of an individual system in the equilibrium ensemble can be described by a trajectory $\alpha(t)$ of the reaction coordinate. A transition path (or reactive trajectory) refers to a segment of the equilibrium trajectory in which a transition occurs. To precisely define the transition path, one needs to specify the boundaries of the transition region, although for rare transitions the kinetic rates are insensitive to the exact definition. For the sake of simplicity, here I use the free energy minima α_A and α_B as the boundaries for the transition path, although other choices should be perfectly valid as well. With such a definition, a transition path from A to B must start at α_A and end at α_B , and all intermediate values of the path must lie between α_A and α_B . A transition path from B to A has similar requirements.

Each unique transition path, denoted as $\text{TP}_{A \rightarrow B}^{(i)}$, is defined by a unique piece of trajectory $\text{TP}_{A \rightarrow B}^{(i)}(t')$ in which the reaction coordinate evolves from α_A to α_B . The time origin of $\text{TP}_{A \rightarrow B}^{(i)}(t')$ can be shifted such that $t' \in [0, t_{\text{TP}}^{(i)}]$, in which $t_{\text{TP}}^{(i)}$ denotes the total duration (length) of the path. The transition path must satisfy $\text{TP}_{A \rightarrow B}^{(i)}(0) = \alpha_A$, $\text{TP}_{A \rightarrow B}^{(i)}(t_{\text{TP}}^{(i)}) = \alpha_B$, and $\alpha_A < \text{TP}_{A \rightarrow B}^{(i)}(t') < \alpha_B$ for $0 < t' < t_{\text{TP}}^{(i)}$. In the equilibrium ensemble, the rate of launching the transition path $\text{TP}_{A \rightarrow B}^{(i)}$ (leaving state A) is denoted by $k^{(i)}$. Obviously, $k^{(i)}$ is also the rate of completing the transition $\text{TP}_{A \rightarrow B}^{(i)}$ (reaching state B). Furthermore, the reverse trajectory of $\text{TP}_{A \rightarrow B}^{(i)}(t')$ defines a transition path $\text{TP}_{B \rightarrow A}^{(i)}$ from B to A with the same length. In the equilibrium ensemble, the rates for $\text{TP}_{A \rightarrow B}^{(i)}$ and $\text{TP}_{B \rightarrow A}^{(i)}$ are both $k^{(i)}$ due to detailed balance. Importantly, the overall transition rate k_0 defined earlier is the probability of starting *any* transition path from A (or, equivalently, completing *any* transition at B) in unit time in the equilibrium ensemble. Therefore, k_0 is the sum of the rates for all unique transition paths,

$$k_0 = \sum_{(i)} k^{(i)}. \quad (3)$$

Note that for the sake of simplicity, I assign a discrete label (i) to each unique transition path, whereas in principle the

distribution of the paths should be continuous. Nonetheless, this distinction does not affect any conclusion in this study.

Now consider an interval $[\alpha_1, \alpha_2]$ in the transition region, with $\alpha_A \leq \alpha_1 < \alpha_2 \leq \alpha_B$. In the following, I will calculate the joint equilibrium probability $P(\text{TP}_{A \rightarrow B}^{(i)}, \alpha_1 \leq \alpha \leq \alpha_2)$ for finding the system on the transition path $\text{TP}_{A \rightarrow B}^{(i)}$ and simultaneously with the reaction coordinate inside the interval $[\alpha_1, \alpha_2]$. If the trajectory $\text{TP}_{A \rightarrow B}^{(i)}(t')$ crosses the interval in a single pass, from t'_1 to t'_2 , visiting the interval at the current time (assumed to be 0) would correspond to starting the transition path $\text{TP}_{A \rightarrow B}^{(i)}$ during $[-t'_2, -t'_1]$, and the probability $P(\text{TP}_{A \rightarrow B}^{(i)}, \alpha_1 \leq \alpha \leq \alpha_2)$ is therefore given by $k^{(i)} \cdot (t'_2 - t'_1)$. In more general cases, $\text{TP}_{A \rightarrow B}^{(i)}$ may visit the interval during multiple periods, and the probability is given below

$$\Delta(\alpha; \alpha_1, \alpha_2) \equiv \begin{cases} 1, & \text{if } \alpha_1 \leq \alpha \leq \alpha_2 \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

$$\tau_{\text{TP}}^{(i)}(\alpha_1, \alpha_2) \equiv \int_0^{t_{\text{TP}}^{(i)}} \Delta(\text{TP}_{A \rightarrow B}^{(i)}(t'); \alpha_1, \alpha_2) dt', \quad (5)$$

$$P(\text{TP}_{A \rightarrow B}^{(i)}, \alpha_1 \leq \alpha \leq \alpha_2) = k^{(i)} \cdot \tau_{\text{TP}}^{(i)}(\alpha_1, \alpha_2), \quad (6)$$

in which $\tau_{\text{TP}}^{(i)}(\alpha_1, \alpha_2)$ is the total duration for $\text{TP}_{A \rightarrow B}^{(i)}(t')$ within the interval $[\alpha_1, \alpha_2]$. The same relation can be alternatively derived by considering the probability and the rate based on a single long equilibrium trajectory.¹⁸

The joint probability above can also be expressed as

$$P(\text{TP}_{A \rightarrow B}^{(i)}, \alpha_1 \leq \alpha \leq \alpha_2) = P(\text{TP}_{A \rightarrow B}^{(i)} | \alpha_1 \leq \alpha \leq \alpha_2) \cdot P(\alpha_1 \leq \alpha \leq \alpha_2), \quad (7)$$

in which $P(\alpha_1 \leq \alpha \leq \alpha_2)$ is the marginal equilibrium probability for $\alpha_1 \leq \alpha \leq \alpha_2$ regardless of whether the microstate is on a transition path or not, and $P(\text{TP}_{A \rightarrow B}^{(i)} | \alpha_1 \leq \alpha \leq \alpha_2)$ is the conditional equilibrium probability for being on the transition path $\text{TP}_{A \rightarrow B}^{(i)}$ given that the reaction coordinate is inside $[\alpha_1, \alpha_2]$. Combining Eqs. (6) and (7), one obtains

$$k^{(i)} = \frac{P(\text{TP}_{A \rightarrow B}^{(i)} | \alpha_1 \leq \alpha \leq \alpha_2) \cdot P(\alpha_1 \leq \alpha \leq \alpha_2)}{\tau_{\text{TP}}^{(i)}(\alpha_1, \alpha_2)}. \quad (8)$$

According to Eq. (3), the overall transition rate k_0 can then be expressed as

$$k_0 = P(\alpha_1 \leq \alpha \leq \alpha_2) \cdot \sum_{(i)} \frac{P(\text{TP}_{A \rightarrow B}^{(i)} | \alpha_1 \leq \alpha \leq \alpha_2)}{\tau_{\text{TP}}^{(i)}(\alpha_1, \alpha_2)}. \quad (9)$$

To evaluate the summation in Eq. (9), I define a quantity $\lambda(\alpha_1, \alpha_2)$ for each microstate in the equilibrium trajectory,

$$\lambda(\alpha_1, \alpha_2) \equiv \begin{cases} 1/\tau_{\text{TP}}^{(i)}(\alpha_1, \alpha_2), & \text{if on } \text{TP}_{A \rightarrow B}^{(i)} \text{ or } \text{TP}_{B \rightarrow A}^{(i)} \\ 0, & \text{if not on any TP} \end{cases}. \quad (10)$$

$\lambda(\alpha_1, \alpha_2)$ thus depends on the history and future of the current microstate in the equilibrium trajectory. The expected value (average) of $\lambda(\alpha_1, \alpha_2)$ for the equilibrium ensemble of

microstates with α inside the interval $[\alpha_1, \alpha_2]$ is then given by

$$\langle \lambda(\alpha_1, \alpha_2) \rangle_{\alpha_1 \leq \alpha \leq \alpha_2} = 2 \sum_{(i)} \frac{P(\text{TP}_{A \rightarrow B}^{(i)} | \alpha_1 \leq \alpha \leq \alpha_2)}{\tau_{\text{TP}}^{(i)}(\alpha_1, \alpha_2)}. \quad (11)$$

The factor of 2 above arises because $\lambda(\alpha_1, \alpha_2)$ counts for the transitions in both directions, $\text{TP}_{A \rightarrow B}^{(i)}$ and $\text{TP}_{B \rightarrow A}^{(i)}$, which have equal probability. In combination with Eq. (9), the rate constant can be determined as

$$k_0 = \frac{1}{2} P(\alpha_1 \leq \alpha \leq \alpha_2) \cdot \langle \lambda(\alpha_1, \alpha_2) \rangle_{\alpha_1 \leq \alpha \leq \alpha_2}. \quad (12)$$

Equation (12) offers an approach to calculate the transition rate k_0 . First, the probability $P(\alpha_1 \leq \alpha \leq \alpha_2)$, a thermodynamic quantity, can be obtained as $P(\alpha_1 \leq \alpha \leq \alpha_2) = \int_{\alpha_1}^{\alpha_2} p_\alpha(\alpha) d\alpha$. To further obtain $\langle \lambda(\alpha_1, \alpha_2) \rangle_{\alpha_1 \leq \alpha \leq \alpha_2}$, one would sample a set of microstates within the interval $[\alpha_1, \alpha_2]$ according to the equilibrium (Boltzmann) distribution, and then evaluate their $\lambda(\alpha_1, \alpha_2)$ values by shooting off trajectories forward and backward in time.^{16–18} Specifically, for each microstate, a forward trajectory and a backward trajectory (by reverting the velocities of all atoms) would be initiated in two unbiased simulations, respectively, each of which will run until the reaction coordinate reaches α_A or α_B . If the forward and backward trajectories commit to the same state, they do not form a transition path and $\lambda(\alpha_1, \alpha_2)$ will be 0. On the other hand, if the two trajectories commit to different states, they constitute a reversible transition path (reactive trajectory) between α_A and α_B , with $\lambda(\alpha_1, \alpha_2)$ determined (Eq. (10)) by its duration $\tau_{\text{TP}}^{(i)}(\alpha_1, \alpha_2)$. The average $\lambda(\alpha_1, \alpha_2)$ among all the examined microstates within $[\alpha_1, \alpha_2]$ will thus be an unbiased estimate for $\langle \lambda(\alpha_1, \alpha_2) \rangle_{\alpha_1 \leq \alpha \leq \alpha_2}$, and the transition rate k_0 can be subsequently calculated from Eq. (12).

The approach above is general and rigorous for any chosen interval $[\alpha_1, \alpha_2]$, given that the examined microstates in this interval are properly drawn from the equilibrium distribution, with α values at lower free energies sampled more frequently. For narrow intervals, Eq. (12) can be written in more familiar terms. Suppose $\alpha_1 = \alpha_0 - \delta\alpha/2$ and $\alpha_2 = \alpha_0 + \delta\alpha/2$, with α_0 and $\delta\alpha$ the center and width of the interval, respectively. If the width $\delta\alpha$ is sufficiently small such that $p_\alpha(\alpha)$ and $G(\alpha)$ can be considered constant within the interval, it holds $P(\alpha_0 - \frac{\delta\alpha}{2} \leq \alpha \leq \alpha_0 + \frac{\delta\alpha}{2}) = p_\alpha(\alpha_0) \cdot \delta\alpha$. Furthermore, the average λ value for the microstates at any α within a sufficiently small interval can also be considered a constant. In such cases, Eq. (12) can be expressed as

$$k_0 = \frac{1}{2} p_\alpha(\alpha_0) \cdot \delta\alpha \cdot \left\langle \lambda(\alpha_0 - \frac{\delta\alpha}{2}, \alpha_0 + \frac{\delta\alpha}{2}) \right\rangle_{\alpha_0}, \quad (13)$$

in which the average $\langle \lambda(\alpha_0 - \frac{\delta\alpha}{2}, \alpha_0 + \frac{\delta\alpha}{2}) \rangle_{\alpha_0}$ is over the microstates near the dividing surface at α_0 , which can be sampled from simulations with the reaction coordinate α strongly restrained at α_0 . According to Eq. (13), the rate k_0 can be calculated at any α_0 . In particular, at the free energy barrier α^* , $p_\alpha(\alpha)$ is at the minimum and the average λ is at the maximum. Choosing α_0 at this position will therefore make the rate calculation more efficient due to a higher probability of sampling transition paths.

The conclusions, Eqs. (12) and (13), in this study are consistent with previous relations for the rate calculation.^{16–18} In particular, each time the transition path crosses an interval of width $\delta\alpha$, the duration is given by $\delta\alpha/|v|$, in which v is the average velocity during the crossing time. When $\delta\alpha$ approaches zero, v will become the instantaneous velocity when the dividing surface at α_0 is crossed. Consequently, the zero-width limit of $\delta\alpha \cdot \lambda(\alpha_0 - \frac{\delta\alpha}{2}, \alpha_0 + \frac{\delta\alpha}{2})$ for a given transition path is equal to $(\sum_j |v_j|^{-1})^{-1}$, summing over the individual crossings (each with an instantaneous velocity v_j) of the dividing surface. Therefore, the rate formula in Ref. 16 can be rigorously derived from Eq. (13) here. Technically, evaluating the durations (as in Eq. (13)) would be slightly more convenient than evaluating the instantaneous crossing speeds. Furthermore, Eq. (12) in this study is essentially the same relation presented in Ref. 18, which is proven here to be general and exact for any interval, but with the caveat that the examined microstates should be properly drawn from the equilibrium distribution. Finally, all approaches above are equivalent¹⁸ to the Bennett-Chandler formalism^{1,14,15} but can be much more efficient than direct calculations of the transmission coefficient.^{1,14} In addition, if all trajectories at the dividing surface are reactive and cross the surface only once, all methods will predict the rate in the transition state theory.^{1,14,16}

III. NUMERICAL SIMULATIONS

I performed calculations on a model system to verify the relations derived here. The system contains a single particle of mass m in a two-dimensional space under the potential

$$U(x, y) = -k_B T \ln \left(\exp \left\{ -\frac{1}{2} \left[\left(\frac{x+b}{l_0} \right)^2 + \left(\frac{y+b}{l_0} \right)^2 \right] \right\} + \exp \left\{ -\frac{1}{2} \left[\left(\frac{x-b}{l_0} \right)^2 + \left(\frac{y-b}{l_0} \right)^2 \right] \right\} \right). \quad (14)$$

I use $k_B T$, l_0 , $v_0 \equiv \sqrt{\frac{k_B T}{m}}$, and $t_0 \equiv l_0/v_0$ as the units for energy, length, velocity, and time, respectively. As shown in Fig. 1(a), the potential energy features two local minima at $(-b, -b)$ and (b, b) , respectively. I set $b = 3.5l_0$ in this test. The motion of the particle is governed by Langevin dynamics at a constant temperature T and a friction coefficient $\gamma = 500 k_B T/(l_0 v_0)$. Under such conditions, the Smoluchowski diffusion equation in the limit of strong friction is valid, predicting a spontaneous transition rate $k_0 = 5.7 \times 10^{-9} t_0^{-1}$, which is used to validate the calculated rates from the MD simulations below.

Two independent sets of calculations (Table I) were performed, each using a different linear reaction coordinate, RC1 or RC2, along the direction of $(\hat{x} + \hat{y})/\sqrt{2}$ or \hat{x} , respectively (Fig. 1(a)). The free energy as a function of each reaction coordinate is shown in Fig. 1(b). For the first set of calculations, a total of 100 microstates with RC1 at the peak ($\alpha^* = 0$) of the free energy were selected. The positions of these microstates were drawn from the Boltzmann distribution along the line (dividing surface) perpendicular to RC1, and the velocities were assigned according to the Maxwell distribution. For each microstate, a pair of unbiased MD simulations (with a time step of $10^{-4} t_0$) was initiated: the forward simulation simply started

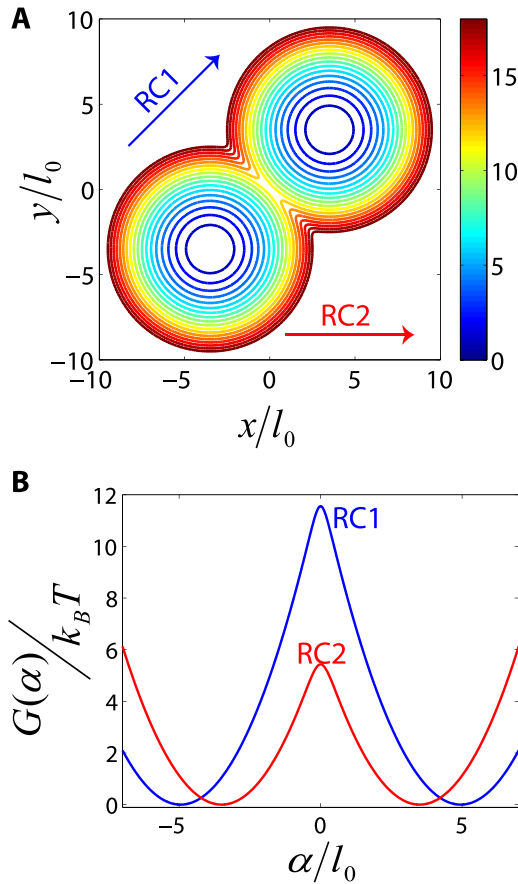


FIG. 1. (a) The contour plot for the potential surface (Eq. (14)) in the model system. The directions of two alternative reaction coordinates (RC1 and RC2) are indicated. (b) The free energy profile along each reaction coordinate.

with the given microstate, and the backward simulation started with the same position but reverted velocity. Each simulation was stopped when the reaction coordinate (RC1) reached either free energy minimum. The second set of calculations with RC2 as the reaction coordinate was done similarly, except with 1000 pairs of forward/backward trajectories initiated at the barrier top.

Each pair of forward/backward trajectories was assigned a λ value (Eq. (10)), which is the inverse of the duration within the given interval if the two trajectories commit to different states (thus forming a transition path), or otherwise zero. The average λ among all the trajectory pairs was then used to

TABLE I. Summary for the rate calculations using the two reaction coordinates (RC1 and RC2), respectively. N_{sim} is the total number of forward/backward simulation pairs performed. N_{TP} is the number of simulation pairs that form a transition path. $p_{\alpha}(0)$ is the equilibrium probability density at the peak of the free energy (see Fig. 1(b)). The value $\delta\alpha \cdot \overline{\lambda}(-\delta\alpha/2, \delta\alpha/2)$ is taken from Fig. 2. The rate constant k_0 is then calculated according to Eq. (13). In comparison, the theoretical transition rate is $k_0 = 5.7 \times 10^{-9} \tau_0^{-1}$.

	N_{sim}	N_{TP}	$p_{\alpha}(0)$ (l_0^{-1})	$\delta\alpha \cdot \overline{\lambda}(-\delta\alpha/2, \delta\alpha/2)$ (v_0)	k_0 (τ_0^{-1})
RC1	100	56	1.91×10^{-6}	$(6.6 \pm 1.0) \times 10^{-3}$	$(6.3 \pm 0.9) \times 10^{-9}$
RC2	1000	2	8.73×10^{-4}	$(9.6 \pm 7.1) \times 10^{-6}$	$(4.2 \pm 3.1) \times 10^{-9}$

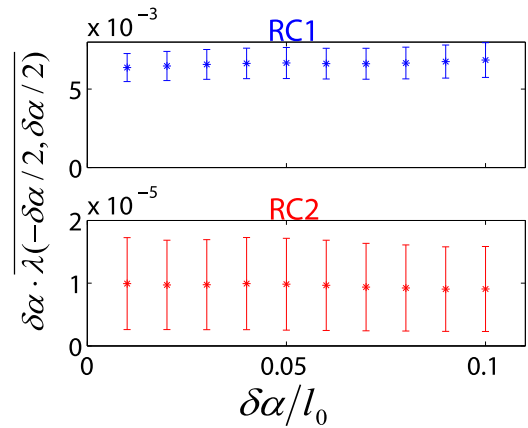


FIG. 2. Analysis of the simulations initiated at the free energy barrier ($\alpha_0 = 0$). For each pair of forward/backward trajectories, the $\lambda(-\frac{\delta\alpha}{2}, \frac{\delta\alpha}{2})$ value is calculated (Eq. (10)) based on whether they form a transition path and the duration of the path in the interval $[-\frac{\delta\alpha}{2}, \frac{\delta\alpha}{2}]$. The average $\overline{\lambda}(-\frac{\delta\alpha}{2}, \frac{\delta\alpha}{2})$ is then calculated over all the trajectory pairs. Intervals of different widths ($\delta\alpha$) are used, and the product $\delta\alpha \cdot \overline{\lambda}(-\frac{\delta\alpha}{2}, \frac{\delta\alpha}{2})$ (in units of v_0) is shown for each interval.

calculate the kinetic rate according to Eq. (13). Although λ clearly depends on the width ($\delta\alpha$) of the interval, Fig. 2 shows that the product $\delta\alpha \cdot \overline{\lambda}(-\frac{\delta\alpha}{2}, \frac{\delta\alpha}{2})$ is a constant and hence the rate can be calculated using any small interval. Furthermore, the calculated rates (Table I) are in good agreement with the theoretical value, thus validating the approach here.

Among the two reaction coordinates in the test, RC1 has a higher free energy barrier (thus a smaller probability at the barrier) and a larger $\overline{\lambda}$ than RC2 does. The rate constants, determined by the product of the probability and $\overline{\lambda}$, are nonetheless similar in both cases, thus confirming that the intrinsic transition rate should not depend on the artificial choice of the reaction coordinate. For RC1, about half of the microstates at the free energy peak are on transition paths (Table I), thus indicating a good reaction coordinate.^{16,17} In contrast, the fraction ($\sim 0.2\%$) is very small for RC2 (Table I), which is thus a poor reaction coordinate, as the rate calculation with RC2 suffers a poor statistical accuracy (Table I) despite employing ten times more simulations than in the RC1 case. The poor quality of RC2 is because a relevant degree of freedom (y) is completely ignored, such that most microstates at the free energy barrier have committers^{5,17} (splitting probabilities) far from 1/2. This example demonstrates that although theoretically the transition rate can be correctly calculated based on any reaction coordinate, adopting a good one could significantly enhance the efficiency. In principle, the Bennett-Chandler approach^{1,14,15} could also calculate the transition rate from the simulations here, but the statistical uncertainty is very high even for the good reaction coordinate (RC1) due to the small (< 0.01) transmission coefficient.

IV. CONCLUSIONS

This study provides a rigorous and practical approach to obtain the transition rates for two-state systems from MD simulations. First, a reaction coordinate α would be chosen and a one-dimensional free energy $G(\alpha)$ be calculated using enhanced sampling techniques,⁶ thus determining the

equilibrium probability distribution $p_\alpha(\alpha)$. Next, a set of microstates near the dividing surface at the free energy barrier α^* will be sampled according to the equilibrium distribution, e.g., by taking snapshots from restrained simulations. Then a forward and a backward simulation will be initiated from each microstate. If the pair of simulations forms a transition path, the duration of the path in a small interval at α^* can be obtained. The transition rate can then be calculated from the probability $p_\alpha(\alpha^*)$ and the durations of the transition paths, according to Eq. (13). As demonstrated here, the approach remains efficient when the barrier crossing is highly diffusive, although it does not explicitly invoke any diffusion model or the approximations therein. Furthermore, the transition paths generated here, after being properly weighted¹⁶ by the inverse of their durations (see Eq. (6)), will quantitatively reveal the detailed mechanism for the rare transitions.

The fraction of transition paths among the examined trajectories indicates the quality of the chosen reaction coordinate.^{16,17} Whereas the approach here is general and valid for any reaction coordinate, a good one will lead to a more efficient sampling of the transition paths and thus a more reliable rate calculation. In contrast, poor reaction coordinates typically do not properly incorporate some slow degrees of freedom, and consequently only a small fraction of the microstates at the dividing surface are on transition paths, thus making the calculation of the kinetic factor inefficient. Furthermore, the free energy calculation and the sampling of the microstates at the dividing surface may also become problematic if some orthogonal degrees of freedom involve slow equilibration. Earlier studies^{10,20} suggested that Hamiltonian replica exchange²¹ in umbrella sampling simulations could facilitate the sampling of the orthogonal coordinates and to some extent alleviate the problems with an imperfect reaction coordinate. Furthermore, from the simulation trajectories, some optimization techniques¹⁷ may be applied to obtain improved reaction coordinates. With the available

theories and computational methods, including the approach presented in this study, the calculation of transition rates from MD simulations should now become feasible and reliable for many systems.

ACKNOWLEDGMENTS

The calculations were performed on a Linux cluster at the School of Science, IUPUI.

- ¹D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic Press, San Diego, 2002).
- ²P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291–318 (2002).
- ³T. S. van Erp and P. G. Bolhuis, *J. Comput. Phys.* **205**(1), 157–181 (2005).
- ⁴T. S. van Erp, *J. Chem. Phys.* **125**(17), 174106 (2006).
- ⁵B. Peters, *Annu. Rev. Phys. Chem.* **67**, 669–690 (2016).
- ⁶R. C. Bernardi, M. C. Melo, and K. Schulten, *Biochim. Biophys. Acta* **1850**(5), 872–877 (2015).
- ⁷B. Peters, P. G. Bolhuis, R. G. Mullen, and J. E. Shea, *J. Chem. Phys.* **138**(5), 054106 (2013).
- ⁸G. Hummer, *New J. Phys.* **7**, 34 (2005).
- ⁹F. Zhu and G. Hummer, *J. Chem. Theory Comput.* **8**(10), 3759–3768 (2012).
- ¹⁰H. Song and F. Zhu, “Finite temperature string method with umbrella sampling: Application on a side chain flipping in Mhp1 transporter,” *J. Phys. Chem. B* (published online).
- ¹¹G. A. Huber and S. Kim, *Biophys. J.* **70**(1), 97–110 (1996).
- ¹²A. K. Faradjian and R. Elber, *J. Chem. Phys.* **120**(23), 10880–10889 (2004).
- ¹³E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **130**(19), 194101 (2009).
- ¹⁴D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).
- ¹⁵D. Chandler, *J. Chem. Phys.* **68**(6), 2959–2970 (1978).
- ¹⁶G. Hummer, *J. Chem. Phys.* **120**(2), 516–523 (2004).
- ¹⁷R. B. Best and G. Hummer, *Proc. Natl. Acad. Sci. U. S. A.* **102**(19), 6732–6737 (2005).
- ¹⁸J. Daru and A. Stirling, *J. Chem. Theory Comput.* **10**(3), 1121–1127 (2014).
- ¹⁹B. Ensing, A. Laio, M. Parrinello, and M. L. Klein, *J. Phys. Chem. B* **109**(14), 6676–6687 (2005).
- ²⁰C. Neale, C. Madill, S. Rauscher, and R. Pomes, *J. Chem. Theory Comput.* **9**(8), 3686–3703 (2013).
- ²¹H. Fukunishi, O. Watanabe, and S. Takada, *J. Chem. Phys.* **116**(20), 9058–9067 (2002).